Description of the Growth Models

Douwe Molenaar[1,4], **Rogier van Berlo**[2,4], **Dick de Ridder**[2,4] **& Bas Teusink**[1,3,4]

1. Centre for Integrative Bioinformatics (IBIVU), Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands
2. Information and Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
3. Top Institute Food and Nutrition, Nieuwe Kanaal 9A, 6709 PA Wageningen, The Netherlands
4. Kluyver Centre for Genomics of Industrial Fermentation, Julianalaan 67, 2628 BC Delft, The Netherlands

We consider a growth medium with an infinite volume and a constant substrate concentration S in which cells are growing. The total biomass of these cells increases exponentially. During balanced growth, in an exponentially growing system, every extensive variable X, like total pools of cellular components, total intracellular volume, total membrane surface etc., increases in time (t) with the same factor (μ) relative to its size:

$$\frac{dX}{dt} = \mu \cdot X \tag{1}$$

The exponential factor μ is called the "specific growth rate". The solution to this differential equation describes exponential growth of a self-replicating system:

$$X(t) = X_{t=0} \cdot e^{\mu \cdot t} \tag{2}$$

From equation (1) it follows that the sum of the rates (v) of synthesis and degradation processes for X are balanced with the growth of X:

$$\frac{dX}{dt} - \mu \cdot X = \sum v_{synthesis} - \sum v_{degradation} - \mu \cdot X = 0 \tag{3}$$

The fact that all extensive variables increase with the same relative growth rate implies that any two extensive variables X_m and X_n will have constant ratios, or in other words that the composition of the cells remains constant. This condition corresponds to the definition of balanced growth (Ingraham et al., 1983):

$$\frac{d(X_m/X_n)}{dt} = \frac{1}{X_n} \cdot \left(\frac{dX_m}{dt} - \frac{X_m}{X_n} \cdot \frac{dX_n}{dt} \right) = \frac{1}{X_n} \cdot (\mu \cdot X_m - \mu \cdot X_m) = 0$$

Since we want to determine the relation between the relative growth rate and the (time-invariant) composition of a cell we need to define a system of equations in terms of the extensive variables relative to a reference extensive variable. The extensive variables

will then transform to intensive variables. The most natural choice for such a reference extensive variable is that of the total intracellular volume Vol. Any extensive variable that refers to a total pool N_X of a compound X will thus transform to the concentration[1] of that compound, written in this document as $c_X = N_X/Vol$. The total intracellular volume of the cells is proportional to the total membrane surface with a proportionality factor β that equals the volume/surface ratio of the particular cell-shape. Here we assume that the membrane is made up of lipid and transporter molecules. N_{lip} and N_{tr} are the total pool sizes of lipid and transporter molecules and A_{lip} and A_{tr} are the specific surfaces of lipid and transporter molecules:

$$Vol = \beta \cdot \left(A_{lip} \cdot N_{lip} + A_{tr} \cdot N_{tr}\right) \tag{4}$$

Expressing this equation relative to the total intracellular volume by dividing both sides of equation (4) by Vol yields:

$$1 = \beta \cdot \left(A_{lip} \cdot c_{lip} + A_{tr} \cdot c_{tr}\right) \tag{5}$$

where c_{lip} and c_{tr} are the ratios of lipid and transporter relative to intracellular volume (N_{lip}/Vol and N_{tr}/Vol, respectively).To illustrate the relation of a shape factor to cellular dimensions: the shape factor of a cylindrically shaped cell with infinite length and a variable diameter r would be:

$$\beta = \frac{r}{2}$$

And for a more realistic cylindrically shaped cell with constant diameter R, hemispherically shaped poles and a variable length d it would be:

$$\beta = \frac{3 \cdot d \cdot R + 4 \cdot R^2}{6 \cdot d + 12 \cdot R}$$

In this case β would vary between $R/3$ at $d = 0$ (coccus-like shape) and $R/2$ at high d (filamentous shape)[2].

The system of equations

The system of equations is stated in terms of intensive variables, relative to the intracellular volume (c_X for each cell component X).

Definition and classification of cellular components

We define all cellular components as the set Cmp. The set of proteins $Prot$ is a subset of Cmp, as is the set of metabolites Met. $Prot$ and Met have no components in common and together form the complete set of cellular components:

$$Prot \cap Met = \emptyset$$
$$Prot \cup Met \equiv Cmp$$

[1] Strictly, we should only speak of the pool/volume ratio since, for example for compounds located in the membrane, this ratio can not be interpreted as the chemical concentration.

[2] Here, a bigger more filamentous cell has a larger surface/volume ratio because a sphere has the lowest surface/volume ratio possible. Filamentation of microorganisms is often observed in response to nutrient limitation.

The set of enzymes Enz is a subset of $Prot$. Enz includes the ribosomes. The set of membrane located components Mem is a subset of Cmp and may contain elements belonging to $Prot$ (like transporters) or Met (like lipids). In the models described here Mem consists only of lipid and transporter proteins.

$$Enz \subset Prot$$
$$Mem \subset Cmp$$

$Memp$ is the set of proteins located in the membrane:

$$Memp \equiv Mem \cap Prot$$

Cell volume

The first equation is a generalization and slight modification of equation (5) and relates the membrane surface to the intracellular volume via the cell shape factor β:

$$1 = \beta \cdot \sum_{i \in Mem} A_i \cdot c_i \tag{I}$$

Mass balance equations

Proteins are synthesized by the ribosome, including the ribosome itself. We model regulation of protein synthesis by assuming that a certain fraction a_j of the ribosomes is occupied with the synthesis of protein j. As for any cell component (3), the synthesis of each protein j has to be in balance with the growth rate. This results in the following system of equations:

$$\alpha_j \cdot v_{rib} - \mu \cdot c_j = 0 \qquad \text{for all } j \in Prot \tag{II}$$

where v_{rib} is the catalytic rate of the total ribosome pool. It also implies that the α_j's are non-negative and add up to one:

$$\alpha_j \geq 0 \qquad \text{for all } j \in Prot \tag{C1}$$
$$\sum_{j \in Prot} \alpha_j = 1 \tag{III}$$

The synthesis and degradation of metabolites can be conveniently written using a notation that includes the reaction stoichiometry matrix (Clarke, 1980). For metabolites as well, the sum of synthesis and degradation rates has to be in balance with the growth rate:

$$\sum_{n \in Enz} a_{mn} \cdot v_n - \mu \cdot c_m = 0 \qquad \text{for all } m \in Met \tag{IV}$$

In the reaction stoichiometry matrix the factors a_{mn} register how much of metabolite m is produced (positive a_{mn}) or consumed (negative a_{mn}) in the elementary reaction catalized by enzyme n.

Additional constraints

Life is subject to physical and biochemical constraints and it often seeks the boundaries of these constraints when optimizing fitness (Koch, 1985). Such constraints are modelled here as mathematical inequalities. To meet the requirements of the claim that our

model derives from first principles, these constraints need to be carefully formulated, i.e. they should derive from basic (bio)physical knowledge. In this model we used three types of constraints.

Positive variables

The regulation factors α_j are positive (see (C1)) because a negative factor would imply that the ribosomes would consume a protein, which is an unlikely event. Also, all concentrations are positive because negative concentrations c_X are not compatible with physical reality.

$$c_X \geq 0 \qquad \text{for all } X \in Cmp \tag{C2}$$

Membrane integrity

The membrane of the described model cell consists of transporter proteins and lipid. In real cells typically half of the membrane consists of membrane proteins (Kadner, 1996), implying that membrane proteins are a substantial factor in determining the intracellular volume. Optimization of the growth rate of the model cell would usually lead to a cell that produces no lipid, because lipid has no enzymatic function but its production requires the consumption of resources that could be employed for synthesis of enzymes. The reason that lipids are produced is because they are important for membrane integrity. Here we mimic this fact by stating that it is necessary to have a membrane protein/lipid ratio of maximally PL_{max}:

$$\frac{\sum_{p \in Memp} c_p}{c_{lip}} \leq PL_{max} \tag{C3}$$

The value chosen in the model simulations below was $PL_{max} = 1$.

Volume occupied by proteins

Proteins occupy a certain volume, which means that the intracellular volume must be minimally equal to this protein volume plus the volume of some water and metabolites. In this model we mimic this fact by stating that the total intracellular protein concentration is less than or equal to a maximal value P_{max}:

$$\sum_{j \in Prot} c_j \leq P_{max} \tag{C4}$$

The value chosen in the model simulations below was $P_{max} = 1$. In the model simulations this constraint is usually "acive"-, meaning that the intracellular volume of cells is packed with proteins to a maximal density, as is observed in real cells (Zimmerman & Trach, 1991). One could, to make a more realistic simulation, impose a similar constraint on the sum of all intracellular metabolites, or for example on the water activity, expressed as a function of all the intracellular components. We only used the simple constraint (C4) in the models presented here.

Numerical simulations and models

Optimizing the growth rate

The system of equations I-IV and C1-C5 explicitly contains the growth rate μ, which is the variable that we want to maximize, also called the "objective variable" in optimization problems. For a given set of extracellular substrate concentration value S_k of compounds k, a given set of values for the ribosomal fractions α_j satisfying equality III, and a given cell shape factor β the remaining equalities I, II and IV always yield a system of n nonlinear equations with n unknowns, where n equals one plus the number of cellular metabolites (including lipids) plus the number of proteins. The unknowns are μ and the concentrations of metabolites and proteins. Under given S_k, α_j and β the system may or may not have a feasible solution, i.e. a solution within the boundaries imposed by constraints C1-C4. The optimization task is to choose the α_j and β in such a way that, for a given set of values S_k, the solution is both in the feasible region, and maximizes μ in that region. This task can be summarized as follows:

maximize μ		(objective variable)
	subject to eqn. I-IV	(equality constraints)
	and to eqn. C1-C4	(inequality constraints)
	given S_k	(extracellular substrate concentration of compound k)

We can choose either to perform this optimization at a constant cell-shape parameter β, or to include β as a variable in the optimization of μ. There is something to say for the latter, because it is observed that cells make considerable adaptations to the volume/surface ratio in response to feeding conditions (Ingraham et al., 1983; Koch, 1996). In the models presented here we allow β to vary from 0 to ∞. In this case, constraint (C4) is of particular importance since in its absence the optimality condition would let β tend to 0, which would lead to numerical instability.

The system of equations is nonlinear in the variables, hence we are dealing with nonlinear optimization problems. These optimization problems were solved using the GAMS modelling and optimization system (version 22.6, `http://www.gams.com`). The trial-licensed KNITRO solver was used to solve the models. To confirm global optimality, these solutions were checked against solutions obtained with the LINDOGlobal solver on the NEOS server (`http://www-neos.mcs.anl.gov/`) using both the default and the multistart options. Parameter perturbation analysis revealed that, qualitatively, results stay the same over a range of values around the actual constants used in the model (see as an example the sensitivity analysis of model B below). Therefore, the conclusions drawn in this work are of a general nature for systems as described here. The models formulated in GAMS language, including a description on how to solve them can be found in the file "suppl_models.zip".

Model A: the basic model

The basic model consists of four proteins (ribosome, transporter, metabolic enzyme, lipid biosynthesis enzyme), and three metabolites (substrate, metabolic intermediate, lipid). A schematic drawing of this model is presented in Figure 3 in the paper. The framework of equations has been described above. The only additional information needed to specify the basic model are the enzyme rate equations, i.e. how the v_n and v_{rib} in equations (IV) and (II) depend on the concentrations of metabolites. We model

these as simple Michaelis-Menten equations. The equations and parameter values are given in Table 1.

Model B: alternative metabolic pathways

In this model we replace the metabolic pathway of the basic model with two alternative metabolic pathways. In fact, the new metabolically efficient pathway MetEf has the same kinetic characteristics as the original pathway. The catalytically efficient pathway CatEf has Michaelis-Menten type kinetics with a higher k_{cat} than MetEf, however, it has a lower yield of precursor with a yield factor γ. The model is summarized in Table 2.

Sensitivity analysis of model B

The results from a sensitivity analysis in which parameters of model B were systematically perturbated are given in Table 3. The analysis shows that the existence of the metabolic shift is relatively independent of the parameter values chosen. Two conditions lead to the disappearance of the metabolic shift, namely a decrease of the Michaelis constant of the MetEf enzyme or an increase of the corresponding parameter of the CatEf enzyme. The second interesting observation is the fact that both upper limit constraints are active, i.e. the actual values P and PL equal the maximum under all conditions. This is also the case when the upper limits P_{max} and PL_{max} are decreased or increased to very high values. The third observation is that although all parameters have an effect on the substrate concentration at which the metabolic shift takes place (S_{shift}), some have very large effects. The (discrete) control coefficient C_{shift} allows a comparison of these parameter effects. It is perhaps no surprise that changes in the kinetics of the alternative metabolic pathways CatEf and MetEf have large effects on S_{shift}, but Table 3 shows that also transporter kinetics ($kcat_{tr}$ and Km_{tr}) has large effects. Interpreted in terms of investments, it suggests that when heavy investments must be made by the cell to gather scarce substrate (low concentrations of S) using an inefficient transporter (i.e. with high Km_{tr} or low $kcat_{tr}$), it tends to be beneficial to maximize the biomass yield by using a metabolically efficient pathway.

Model C: alternative metabolic ATP-generating pathways

In this model we replace the metabolic pathway of the basic model with two alternative ATP generating pathways, and we add an intermediate metabolite that need to be "activated" by ATP to become the biomass precursor. As in model B, the MetEf pathway is metabolically efficient, generating one ATP per substrate molecule. The CatEf pathway is metabolically inefficient, only generating $\gamma < 1$ ATP per substrate, but has a higher k_{cat} than the MetEf pathway. The model is summarized in Table 4.

Model D: growth on two substrates

To investigate the effect of one limiting substrate on the efficiency of metabolism of another substrate, we construct a model of a cell growing on two substrates. The basic model was left intact and an additional N-substrate was introduced (to indicate that one could think of a nitrogen source as the second substrate). This N-substrate is imported in the cell by a second transporter (N-transporter) and in a reaction catalyzed by the

ribosome combines with the metabolite to form the building blocks of the proteins. The model is summarized in Table 5.

Model E: the production of recombinant protein

The production of recombinant protein is simulated by letting the cell produce a certain pre-determined fraction (α_{dummy}) of "dummy" protein, i.e. a protein that has no other function than that a fraction of ribosomes is involved in its synthesis and that it occupies cell volume. The α_{dummy} is, of course, unlike the other α_j's a constant parameter and not a variable in optimization. Consequently, to simulate recombinant protein production equation (III) in model B is modified as follows:

$$\sum_{j \in Prot} \alpha_j = 1 - \alpha_{dummy} \qquad \text{(IIIa)}$$

The rest of the model is exactly the same as model B.

References

Clarke B (1980) Stability of complex reaction networks. In Prigogine I, Rice S, eds., *Advances in Chemical Physics*, volume 43, pp. 1–216. Wiley, New York

Ingraham JL, Maaløe O, Neidhardt FC (1983) *Growth of the bacterial cell*. Sinauer Associates, Inc., Sunderland, MA, USA

Kadner RJ (1996) Cytoplasmic membrane. In Neidhardt FC, ed., *Escherichia coli and Salmonella. Cellular and molecular biology*, volume 1, 2^{nd} edition, chapter 7, pp. 58–87. ASM Press, Washington, DC, USA

Koch AL (1985) The macroeconomics of bacterial growth. In Fletcher M, Floodgate GD, eds., *Bacteria in their natural environments.*, chapter 1, pp. 1–42. Academic Press, London, UK

Koch AL (1996) What size should a bacterium be? A question of scale. *Annu Rev Microbiol* **50**: 317–348

Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* **222**: 599–620

Table 1: Components, reactions and rate equations of the basic model

Component	**Symbol, reaction**	**Equations**	**Parameters**
Extracellular substrate	S		
Intracellular substrate	Si		
Precursor	P		
Lipid	Lip		
Transporter	$S \longrightarrow Si$	$v_{tr} = \frac{kcat_{tr} \cdot c_{tr} \cdot [S]}{Km_{tr} + [S]}$	$kcat_{tr}$ =7
			Km_{tr} =1
Metabolic enzyme	$Si \longrightarrow P$	$v_{met} = \frac{kcat_{met} \cdot c_{met} \cdot [Si]}{Km_{met} + [Si]}$	$kcat_{met}$=5
			Km_{met} =1
Ribosome	$P \longrightarrow j \in Prot$	$v_{rib} = \frac{kcat_{rib} \cdot c_{rib} \cdot [P]}{Km_{rib} + [P]}$	$kcat_{rib}$ =3
			Km_{rib} =1
Lipid biosynthesis enzyme	$P \longrightarrow Lip$	$v_{lpb} = \frac{kcat_{lpb} \cdot c_{lpb} \cdot [P]}{Km_{lpb} + [P]}$	$kcat_{lpb}$=5
			Km_{lpb} =1

Table 2: Components, reactions and rate equations of the model with alternative metabolic pathways[†]

Component	**Symbol, reaction**	**Equations**	**Parameters**
MetEf and CatEf replace "Metabolic enzyme" in Table 1			
MetEf pathway	$Si \longrightarrow P$	$v_{metef} = \frac{kcat_{metef} \cdot c_{met} \cdot [Si]}{Km_{metef} + [Si]}$	$kcat_{metef}$=5
			Km_{metef} =1
CatEf pathway	$Si \longrightarrow \gamma P$	$v_{catef} = \frac{kcat_{catef} \cdot c_{catef} \cdot [Si]}{Km_{catef} + [Si]}$	$kcat_{catef}$ =10
			Km_{catef} =1
			γ =0.8

[†] Only changes relative to the basic model (Table 1) are listed

Table 3: Sensitivity analysis[1] of the the model with alternative metabolic pathways (see Table 2)

$\mathbf{kcat_{tr}}$	$\mathbf{kcat_{rib}}$	$\mathbf{kcat_{metef}}$	$\mathbf{kcat_{catef}}$	$\mathbf{kcat_{pb}}$	$\mathbf{Km_{tr}}$	$\mathbf{Km_{rib}}$	$\mathbf{Km_{metef}}$	$\mathbf{Km_{catef}}$	$\mathbf{Km_{pb}}$	$\mathbf{P_{max}}$	$\mathbf{PL_{max}}$	μ_{100}[2]	$\mathbf{S}_{shift}$[3]	$\mathbf{C_{shift}}$[4]	$avg\mathbf{P}$[5]	$avg\mathbf{PL}$[6]
7	3	5	10	5	1	1	1	1	1	1	1	0.587	0.268	*na*[7]	1.0	1.0
3.5												0.452	0.801	-1.6	1.0	1.0
14												0.714	0.108	-1.3	1.0	1.0
	1.5											0.440	0.268	0.0	1.0	1.0
	6											0.718	0.322	0.3	1.0	1.0
		2.5	5									0.477	0.129	1.1	1.0	1.0
		10	20									0.680	0.801	1.6	1.0	1.0
				2.5								0.552	0.322	-0.3	1.0	1.0
				10								0.608	0.268	0.0	1.0	1.0
					0.2							0.588	0.052	1.0	1.0	1.0
					5							0.579	1.660	1.1	1.0	1.0
						0.2						0.699	0.386	-0.2	1.0	1.0
						5						0.418	0.223	-0.1	1.0	1.0
							0.2					0.612	*no shift*	*na*	1.0	1.0
							5					0.587	0.036	-1.2	1.0	1.0
								0.2				0.657	0.108	0.6	1.0	1.0
								5				0.534	*no shift*	*na*	1.0	1.0
									0.2			0.602	0.268	0.0	1.0	1.0
									5			0.531	0.322	0.1	1.0	1.0
										0.5		0.474	0.186	0.5	0.5	1.0
										2		0.702	0.463	0.8	2.0	1.0
										20		1.018	1.384	0.5	20	1.0
										200		1.178	3.443	0.4	200	1.0
											0.5	0.517	0.556	-1.1	1.0	0.5
											2	0.632	0.223	-0.3	1.0	2.0
											20	0.683	0.155	-0.0	1.0	20
											200	0.688	0.129	-0.0	1.0	200

[1]The first line shows the reference parameter settings and results from the original model (Table 2). Subsequent lines show simulations with perturbated parameters. Only the perturbated parameter values are displayed. Note that the k_{cat}'s of the metabolic enzymes CatEf and MetEf were changed concomittantly to guarantee the condition that one route is catalytically efficient while the other is metabolically efficient.

[2]μ_{100}: The growth rate at substrate concentration $S = 100$. This value is close to the maximal growth rate.

[3]S_{shift}: Calculated as the average substrate concentration of the substrate concentrations flanking the stepwise metabolic shift.

[4]C_{shift}: The discrete control coefficient of the perturbed parameter p on S_{shift}, calculated as $\Delta log(S_{shift})/\Delta log(p)$, where the changes were calculated relative to the reference values in the first row, or in case of parameter values 20 and 200 for P_{max} and PL_{max} relative to the S_{shift} values corresponding to parameter values 2 and 20, respectively. Negative values of C_{shift} indicate a decrease of S_{shift} with an increase of the parameter value.

[5]avg$P$: The average of the sum of intracellular protein concentrations over the range of 55 substrate concentration values used in one simulation.

[6]avgPL: The average of the membrane protein/lipid ratios over the range of 55 substrate concentration values used in one simulation.

[7]*na*: not applicable

Table 4: Components, reactions and rate equations of the model with alternative ATP generating metabolic pathways[†]

Component	Symbol, reaction	Equations	Parameters	
MetEf and CatEf replace "Metabolic enzyme" in Table 1				
Intermediate	M			
Energy cofactor	ATP, ADP	$[ADP]+[ATP]=AXP$	AXP	=1
MetEf pathway	$Si+ADP \longrightarrow M+ATP$	$v_{metef}=\frac{kcat_{metef}\cdot c_{metef}\cdot c_{Si}\cdot[ADP]}{[Si]\cdot[ADP]+[Si]\cdot KmADP_{metef}+[ADP]\cdot Km_{metef}}$	$kcat_{metef}$	=5
			Km_{metef}	=1
			$KmADP_{metef}$	=0.5
CatEf pathway	$Si+ADP \longrightarrow M+ATP$	$v_{catef}=\frac{kcat_{catef}\cdot c_{catef}\cdot c_{Si}\cdot[ADP]}{[Si]\cdot[ADP]+[Si]\cdot KmADP_{catef}+[ADP]\cdot Km_{catef}}$	$kcat_{catef}$	=10
			Km_{catef}	=1
			$KmADP_{metef}$	=0.5
Activating enzyme	$M+ATP \longrightarrow P+ADP$	$v_{act}=\frac{kcat_{act}\cdot c_{act}\cdot[M]\cdot[ATP]}{[M]\cdot[ATP]+Km_{prc}\cdot[ATP]+KmATP_{prc}\cdot[M]}$	$kcat_{act}$	=1
			Km_{act}	=1
			$KmATP_{act}$	=1

[†] Only changes relative to the basic model (Table 1) are listed

Table 5: Components, reactions and rate equations of the two-substrate model[†]

Component	**Symbol, reaction**	**Equations**	**Parameters**
Extracellular N-substrate	N		
Intracellular N-substrate	Ni		
N-Transporter	$N \longrightarrow Ni$	$v_{trN} = \frac{kcat_{trN} \cdot c_{trN} \cdot [N]}{Km_{trN} + [N]}$	$kcat_{trN}$ =1
			Km_{trN} =1
Ribosome	$P + Ni \longrightarrow j \in Prot$	$v_{rib} = \frac{kcat_{rib} \cdot c_{rib} \cdot [P] \cdot [Ni]}{[P] \cdot [Ni] + KmN_{rib} \cdot [P] + Km_{rib} \cdot [Ni]}$	$kcat_{rib}$ =1
			Km_{rib} =1
			KmN_{rib}=1

[†] Only changes relative to the basic model (Table 1) are listed